

1 CLAIMS

2 What Is Claimed Is:

3 1. A processing apparatus comprising:

4 structured/hierarchical content, said apparatus makes
5 a determination whether or not said structured/hierarchical
6 content delivered through a network includes a content
7 portion matched with a predetermined matching pattern, and
8 performs predetermined processing for the
9 structured/hierarchical content if a result of the
10 determination is positive, the processing apparatus further
11 comprising:

12 target subtree setting means for setting a target
13 subtree relating to a range including a target content
14 portion as an extracted portion of the matching pattern in
15 the structured/hierarchical content (hereinafter, referred
16 to as a "target content") from which the matching pattern is
17 to be extracted;

18 occurrence mode detecting means for detecting an
19 occurrence mode of each node of the target subtree by
20 selecting a plurality of past structured/hierarchical

1 contents with respect to the target content and collating
2 the target subtree relating to the target content with a
3 tree relating to each of the past structured/hierarchical
4 contents;

5 statistical information generating means for
6 generating statistical information concerning an occurrence
7 frequency of the occurrence mode of each node in the target
8 subtree based on the plurality of past
9 structured/hierarchical contents;

10 classifying means for performing classification of
11 each node of the target subtree based on the statistical
12 information and a result of detecting the occurrence mode;
13 and

14 matching pattern generating means for generating the
15 matching pattern for the target content portion based on the
16 classification.

17 2. The processing apparatus according to claim 1, wherein
18 the predetermined processing is to associate related
19 information with the content portion of the
20 structured/hierarchical content.

1 3. The processing apparatus according to claim 2, wherein
2 the related information includes an annotation.

3 4. The processing apparatus for a structured/hierarchical
4 content according to claim 1, wherein the predetermined
5 processing is processing for copying the content portion of
6 the structured/hierarchical content for a purpose of
7 utilizing the content portion of the structured/hierarchical
8 content for another structured/hierarchical content.

9 5. The processing apparatus for a structured/hierarchical
10 content according to claim 1, wherein the
11 structured/hierarchical content is a Web content.

12 6. The processing apparatus for a structured/hierarchical
13 content according to claim 1, wherein the classifying means
14 classifies nodes of the target subtree into stationary
15 nodes, updated nodes and additional nodes.

16 7. The processing apparatus according to claim 6, wherein
17 the occurrence mode detecting means includes, as the
18 occurrence mode to be detected, (N1) an occurrence mode

1 where detected nodes occur in both of the target content
2 portion and structured/hierarchical contents collated
3 therewith and contents thereof are mutually identical, and
4 (N2) an occurrence mode where the detected nodes occur in
5 both of the target content portion and the
6 structured/hierarchical contents collated therewith and the
7 contents thereof are mutually different, and
8 the classifying means classifies, into the stationary
9 nodes, nodes of which occurrence frequency of the occurrence
10 mode (N1) is determined to be equal to/more than a first
11 threshold value by the statistical information, classifies,
12 into the updated nodes, nodes of which occurrence frequency
13 of the occurrence mode (N2) is determined to be equal
14 to/more than a second threshold value by the statistical
15 information, and classifies, into the additional nodes,
16 nodes other than the stationary nodes and the updated nodes.

17 8. The processing apparatus according to claim 6, wherein
18 the matching pattern generating means includes:

19 repeated portion detecting means for detecting a
20 repeated portion in the target subtree based on the
21 classification into the stationary nodes, the updated nodes

1 and the additional nodes; and
2 repeated information-added matching pattern generating
3 means for generating the matching pattern including presence
4 information of the repeated portion.

5 9. The processing apparatus according to claim 8, wherein
6 the classifying means includes:

7 formed-for-spacer image detecting means for detecting
8 whether or not a node relating to an image is a node
9 relating to a formed-for-spacer image for ensuring a blank
10 region;

11 bullet image detecting means for detecting whether or
12 not the node relating to the image is a node relating to a
13 plurality of bullet images used repeatedly in a same size;

14 first classifying means for classifying the node
15 relating to the formed-for-spacer image into the additional
16 nodes; and

17 second classifying means for allocating a plurality of
18 the nodes relating to the bullet image into a same
19 classification among classifications of the stationary
20 nodes, updated nodes and additional nodes even if display
21 contents of the plurality of nodes are mutually different.

1 10. The processing apparatus according to claim 1, further
2 comprising:
3 collating means for collating the target subtree
4 relating to the target content with the trees relating to a
5 plurality of structured/hierarchical contents adjacent to
6 the target content by selecting the adjacent
7 structured/hierarchical contents in place of the past
8 structured/hierarchical contents with respect to the target
9 content when the past structured/hierarchical contents are
10 not present.

11 11. A processing apparatus comprising:
12 structured/hierarchical content, said apparatus makes
13 a determination whether or not said structured/hierarchical
14 content delivered through a network includes a content
15 portion matched with a predetermined matching pattern, and
16 performs predetermined processing for the
17 structured/hierarchical content if a result of the
18 determination is positive, the processing apparatus
19 comprising:
20 target subtree setting means for setting a target

1 subtree relating to a range including a target content
2 portion as an extracted portion of the matching pattern in
3 the structured/hierarchical content (hereinafter, referred
4 to as a "target content") from which the matching pattern is
5 to be extracted;

6 occurrence mode detecting means for detecting an
7 occurrence mode of each node of the target subtree by
8 selecting a plurality of structured/hierarchical contents
9 adjacent to the target content and collating the target
10 subtree relating to the target content with a tree relating
11 to each of the adjacent structured/hierarchical contents;

12 statistical information generating means for
13 generating statistical information concerning an occurrence
14 frequency of the occurrence mode of each node in the target
15 subtree based on the plurality of adjacent
16 structured/hierarchical contents;

17 classifying means for performing classification of
18 each node of the target subtree based on the statistical
19 information and a result of detecting the occurrence mode;
20 and

21 matching pattern generating means for generating the
22 matching pattern for the target content portion based on the

1 classification.

2 12. A processing method comprising:

3 making a determination whether or not

4 structured/hierarchical content delivered through a network

5 includes a content portion matched with a predetermined

6 matching pattern, and performs predetermined processing for

7 the structured/hierarchical content if a result of the

8 determination is positive, the processing method further

9 comprising:

10 a target subtree setting step of setting a target

11 subtree relating to a range including a target content

12 portion as an extracted portion of the matching pattern in

13 the structured/hierarchical content (hereinafter, referred

14 to as a "target content") from which the matching pattern is

15 to be extracted;

16 an occurrence mode detecting step of detecting an

17 occurrence mode of each node of the target subtree by

18 selecting a plurality of past structured/hierarchical

19 contents with respect to the target content and collating

20 the target subtree relating to the target content with a

21 tree relating to each of the past structured/hierarchical

1 contents;
2 a statistical information generating step of
3 generating statistical information concerning an occurrence
4 frequency of the occurrence mode of each node in the target
5 subtree based on the plurality of past
6 structured/hierarchical contents;

7 a classifying step of performing classification of
8 each node of the target subtree based on the statistical
9 information and a result of detecting the occurrence mode;
10 and

11 a matching pattern generating step of generating the
12 matching pattern for the target content portion based on the
13 classification.

14 13. The processing method according to claim 12, wherein the
15 predetermined processing is to associate related information
16 with the content portion of the structured/hierarchical
17 content.

18 14. The processing method according to claim 13, wherein the
19 related information includes an annotation.

1 15. The processing method according to claim 12, wherein the
2 predetermined processing is processing for copying the
3 content portion of the structured/hierarchical content for a
4 purpose of utilizing the content portion of the
5 structured/hierarchical content for another
6 structured/hierarchical content.

7 16. The processing method according to claim 12, wherein the
8 structured/hierarchical content is a Web content.

9 17. The processing method according to claim 12, wherein in
10 the classifying step, nodes of the target subtree are
11 classified into stationary nodes, updated nodes and
12 additional nodes.

13 18. The processing method according to claim 17, wherein the
14 occurrence mode detecting step includes, as the occurrence
15 mode to be detected, (N1) an occurrence mode where detected
16 nodes occur in both of the target content portion and
17 structured/hierarchical contents collated therewith and
18 contents thereof are mutually identical, and (N2) an
19 occurrence mode where the detected nodes occur in both of

1 the target content portion and the structured/hierarchical
2 contents collated therewith and the contents thereof are
3 mutually different, and
4 in the classifying step, are classified into the
5 stationary nodes, nodes of which occurrence frequency of the
6 occurrence mode (N1) is determined to be equal to/more than
7 a first threshold value by the statistical information, are
8 classified into the updated nodes, nodes of which occurrence
9 frequency of the occurrence mode (N2) is determined to be
10 equal to/more than a second threshold value by the
11 statistical information, and are classified into the
12 additional nodes, nodes other than the stationary nodes and
13 the updated nodes.

14 19. The processing method according to claim 17, wherein the
15 matching pattern generating step includes:

16 a repeated portion detecting step of detecting a
17 repeated portion in the target subtree based on the
18 classification into the stationary nodes, the updated nodes
19 and the additional nodes; and

20 a repeated information-added matching pattern
21 generating step of generating the matching pattern including

1 presence information of the repeated portion.

2 20. The processing method according to claim 19, wherein the
3 classifying step includes:

4 a formed-for-spacer image detecting step of detecting
5 whether or not a node relating to an image is a node
6 relating to a formed-for-spacer image for ensuring a blank
7 region;

8 a bullet image detecting step of detecting whether or
9 not the node relating to the image is a node relating to a
10 plurality of bullet images used repeatedly in a same size;

11 a first classifying step of classifying the node
12 relating to the formed-for-spacer image into the additional
13 nodes; and

14 a second classifying step of allocating a plurality of
15 the nodes relating to the bullet image into a same
16 classification among classifications of the stationary
17 nodes, updated nodes and additional nodes even if display
18 contents of the plurality of nodes are mutually different.

19 21. The processing method according to claim 12, further
20 comprising:

1 a collating step of collating the target subtree
2 relating to the target content with the trees relating to a
3 plurality of structured/hierarchical contents adjacent to
4 the target content by selecting the adjacent
5 structured/hierarchical contents in place of the past
6 structured/hierarchical contents with respect to the target
7 content when the past structured/hierarchical contents are
8 not present.

9 22. A processing method comprising:

10 making a determination whether or not
11 structured/hierarchical content delivered through a network
12 includes a content portion matched with a predetermined
13 matching pattern, and performs predetermined processing for
14 the structured/hierarchical content if a result of the
15 determination is positive, the processing method comprising:
16 a target subtree setting step of setting a target
17 subtree relating to a range including a target content
18 portion as an extracted portion of the matching pattern in
19 the structured/hierarchical content (hereinafter, referred
20 to as a "target content") from which the matching pattern is
21 to be extracted;

1 an occurrence mode detecting step of detecting an
2 occurrence mode of each node of the target subtree by
3 selecting a plurality of structured/hierarchical contents
4 adjacent to the target content and collating the target
5 subtree relating to the target content with a tree relating
6 to each of the adjacent structured/hierarchical contents;

7 a statistical information generating step of
8 generating statistical information concerning an occurrence
9 frequency of the occurrence mode of each node in the target
10 subtree based on the plurality of adjacent
11 structured/hierarchical contents;

12 a classifying step of performing classification of
13 each node of the target subtree based on the statistical
14 information and a result of detecting the occurrence mode;
15 and

16 a matching pattern generating step of generating the
17 matching pattern for the target content portion based on the
18 classification.

19 23. A computer program product comprising a computer usable
20 medium having computer readable program code means embodied
21 therein for causing functions of the apparatus, the computer

1 readable program code means in said computer program product
2 comprising computer readable program code means for causing
3 a computer to effect the functions of claim 1.

4 24. A computer program product comprising a computer usable
5 medium having computer readable program code means embodied
6 therein for causing functions of the apparatus, the computer
7 readable program code means in said computer program product
8 comprising computer readable program code means for causing
9 a computer to effect the functions of claim 11.

10 25. An article of manufacture comprising a computer usable
11 medium having computer readable program code means embodied
12 therein for causing processing, the computer readable
13 program code means in said article of manufacture comprising
14 computer readable program code means for causing a computer
15 to effect the steps of claim 12.

16 26. A program storage device readable by machine, tangibly
17 embodying a program of instructions executable by the
18 machine to perform method steps for processing, said method
19 steps comprising the steps of claim 12.

1 27. An article of manufacture comprising a computer usable
2 medium having computer readable program code means embodied
3 therein for causing processing, the computer readable
4 program code means in said article of manufacture comprising
5 computer readable program code means for causing a computer
6 to effect the steps of claim 22.

7 28. A program storage device readable by machine, tangibly
8 embodying a program of instructions executable by the
9 machine to perform method steps for processing, said method
10 steps comprising the steps of claim 22.